Feasibility, reliability and sensitivity to change of four radiographic scoring methods in patients with psoriatic arthritis

W. Tillett[1] MBChB, BSc, MRCP

D. Jadon[1] MBBCh, MRCP

G. Shaddick[2] MSc, PhD

G. Robinson[3] MBBS, FRCR

R. Sengupta[1] MBBS, MRCP

E. Korendowych[1], PhD, MRCP

C.S. de Vries[4], MA, MSc, PhD, FISPE, FRSM

N.J. McHugh[1,4] MB ChB, MD, FRCP

Royal National Hospital for Rheumatic Diseases, Bath, UK[1]

Department of Mathematics, University of Bath, Bath, UK[2]

Department of Radiology, Royal United Hospital, Bath, UK[3]

Department of Pharmacy and Bath Pharmacology, University of Bath, UK[4]

w.tillett@nhs.net

tel: +44 (0)1225448444

Mobile: +44 (0)7980960722

Fax: +44 (0)1225336809

MeSH terms: Psoriatic Arthritis, Radiography, Outcome assessment

Short title: Radiographic assessment of Psoriatic Arthritis

Word count: 3233 (of which the abstract is 249)

Funding and Disclosures: LOPAS I is funded through an unrestricted grant from Abbvie laboratories Ltd. WT, DJ, GS, GR, CdV declare no conflicts of interest. EK has received honoraria from Pfizer, Abbvie and UCB, NJM has received honoraria from Pfizer and Abbvie.

Objective

We set out to assess the feasibility, reliability and sensitivity to change of four radiographic scoring methods in psoriatic arthritis (PsA).

Methods

Hand and feet radiographs from fifty patients with PsA were scored at two time points with each of the modified Steinbrocker score (STB), modified Sharp score (MSS), Sharp-van der Heijde modified method (VDH) and psoriatic arthritis Ratingen score (PARS) methods by two assessors. The radiographs of ten patients were scored by both assessors to assess reliability using intra-class correlation coefficients (ICC). Sensitivity to change was estimated using a Standardised Response Mean (SRM) and Smallest Detectable Change (SDC).

Results

The patients' mean age at baseline was 50 years (sd 12.1), mean disease duration 10 years (sd 8.4) and mean follow up 25 months (sd 9.6). Intra-rater reliability was excellent for all methods (ICC >0.97). Inter-rater reliability was highest for the VDH (ICC 0.95 to 0.99). The percentage smallest detectable change (SDC) for the STB, PARS, MSS and VDH methods was 2.9%, 2.1%, 1.4% and 1.2%, respectively and the SRMs were 0.46, 0.44, 0.77 and 0.79 respectively. The mean time to score each of the STB, PARS, MSS and VDH methods was 6.2, 10.5, 14.6 and 14.4 minutes, respectively.

Conclusions

The VDH method was the most reliable and sensitive to change but took longer to perform. The STB is the most feasible but lacks the sensitivity of the VDH. The SDC of the PARS is close to that of the VDH and MSS but is quicker to perform.

Significance and Innovations

- Longitudinal observational studies of psoriatic arthritis rarely report radiographic data, in part due to the perceived, but not proven, unfeasibility of existing scores.

- We report the first comparison of feasibility, reliability and sensitivity to change of four radiographic scoring methods for psoriatic arthritis in an observational cohort.

- We have shown that none of the existing radiographic measures are both sufficiently feasible and sensitive to change to be easily applied in large longitudinal observational studies.

- This study highlights the need for existing scores to be modified to a tool that encourages use and prompts collaboration by the psoriatic arthritis research community.

The measurement of radiographic joint damage is essential in characterising disease severity, progression and prognosis. Radiographic damage has been demonstrated in psoriatic arthritis in both early and established disease [1, 2]. It is a core outcome measure in both randomised control trials for novel therapies as well as longitudinal observational studies and is included in the research agenda as a domain of interest by the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) consortium [3].

Several scoring methods have been proposed for use in psoriatic arthritis including the Modified Sharp score (MSS)[4, 5], the Sharp-van der Heijde modified method (VDH)[5, 6], Modified Steinbrocker (STB)[7] and psoriatic arthritis Ratingen score (PARS)[8]. With the exception of the Ratingen method these scoring methods were designed and validated for use in rheumatoid arthritis and subsequently modified for use in psoriatic arthritis.

The choice of radiographic outcome measure to use in psoriatic arthritis randomised controlled trials and longitudinal observational studies was discussed at the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) annual meeting 2012 in Stockholm. There was consensus that the Sharp-van der Heijde was the optimal tool to use in randomised controlled trails (where sensitivity to change is often the most important attribute of the outcome measure) but the most appropriate tool for use in longitudinal observational studies is yet to be determined. Agreement on the use of a single measure in such observational studies would improve comparison of results between cohorts, pooling data and potentially aid meta-analyses. We set out to assess the feasibility, reliability and sensitivity to change of

four radiographic scoring methods in psoriatic arthritis in order to inform discussion on the optimal method for longitudinal observational studies.

Materials and Methods

Postero-anterior radiographs of the hands and feet from fifty consecutive patients commenced on anti-Tumour Necrosis Factor (anti-TNF) therapy for psoriatic arthritis at the Royal National Hospital for Rheumatic Diseases were included. Radiographs taken at the point of anti-TNF commencement and two years prior were scored with each of the modified Steinbrocker score, modified Sharp score, Sharp-van der Heijde method and Ratingen methods. This selection of participants and radiographs was designed to capture patients likely to have sustained active disease and thus progression of radiographic damage upon which the sensitivity to change of each method could be compared. All radiographs included a 'phantom phalanx' as a reference for normal bone density. All selected patient's fulfilled the Classification Criteria for Psoriatic Arthritis (CASPAR) criteria after retrospective assessment of case records[9] [10].

Radiographic scoring methods

Standard antero-posterior radiographs of the hands and forefeet were taken on to *Kodak DirectView* phosphor storage plates. Directly after the images were captured the phosphor plates were scanned and digitised using the *Kodak DirectView* CR 950 system. Images were then viewed and scored on the *Centricity Web viewer* (V3.0.10) on a standard *Hewlett Packard* monitor (1440-900 pixel spacing) with the images preserved at the original 1:1 ratio, all other viewing tools were allowed.

The radiographic techniques are described briefly below and summarised in Table 1. The Steinbrocker is a global technique that scores the joints of the hands and feet for soft tissue swelling or periarticular osteopenia, erosion, joint space narrowing, total destruction (lysis or ankylosis) on a single scale. The Steinbrocker was originally developed for use in rheumatoid arthritis and was modified for use in psoriatic arthritis through inclusion of the distal interphalangeal (DIP) joints.[7] The Sharp-van der Heijde and modified Sharp scores are composite methods that score erosion and joint space narrowing separately and then are summed together in a total score. The Sharp-van der Heijde method was originally developed for use in rheumatoid arthritis then modified for use in psoriatic arhritis [5, 6]. The modified Sharp score was again adapted from the original method used in rheumatoid arthritis [4]. The method we have used in this study is the same as that described by Ravindran *et al.* with the addition that we have included the feet as originally described by van der Heijde [1, 11]. Finally the psoriatic arthritis Ratingen score is the only method to be developed specifically for use in psoriatic arthritis [8]. This composite method scores erosion and bony proliferation characteristic for psoriatic arthritis which are then summed in a total score.

Reading strategy

Two readers, WT and DJ trained in the four scoring methods. This process involved pre-study training in the precise definitions of radiographic findings of psoriatic arthritis as described by Taylor *et al.* [12], literature review of each method, contact with the original authors for clarification where required, then practice with supervision and discussion with an experienced musculoskeletal radiologist (GR) over a two month period. To determine reliability, ten sets of hand and feet radiographs

were scored by both WT and DJ with all four techniques in random order to assess inter-rater reliability, and then scored one month later to estimate intra-rater reliability. The remaining 40 radiographs were then scored (20 by WT and 20 by DJ) with the prior score and radiographs available to optimise sensitivity to change. An assumption of progression only was made for all scores, thus no improvement was recorded.

Statistical analysis

Demographic data was analysed using descriptive statistics. Feasibility was estimated using the average time taken to score each method. Measurement error was estimated by re-scoring ten films by the same rater (intra-rater reliability) and by the other rater (inter-rater reliability). Differences are reported as recommended using both intra-class correlation coefficients (ICC) and visually by plotting the difference in change of scores against the mean change by both raters (Bland-Altman plots) [13].

Sensitivity to change is reported using multiple methods to allow comparison with prior reports. A two-way analysis of variance was performed with an interaction between patient and time leaving a residual from which the standard error of the mean could be estimated. The standardised response mean (SRM) is a unit-less expression of change calculated as the ratio of the mean difference between baseline and follow up score divided by the standard deviation of this difference. A standardised response mean of >0.8 is considered to have a high potential of detecting change. We also report the Smallest Detectable Change (SDC) defined as the smallest difference that can be detected over and above measurement error [14]. A similar method of reporting this is the Smallest Detectable Difference (SDD). The smallest detectable

difference is a less appropriate method to use in this study because we have assessed radiographs with the prior film and score available meaning the assessments cannot be considered to be truly independent. In this instance the smallest detectable difference will overestimate the measurement error, however, we have reported this estimate to allow for comparison with prior reports. We also report both the smallest detectable change and smallest detectable difference a percentage of the total score to allow comparison between methods.

The study was approved by the Bath Research Ethics Committee and has been conducted in accordance with the Declaration of Helsinki. All participants signed informed consent.

Results

All patients fulfilled CASPAR criteria for psoriatic arthritis. The mean age of patients at the baseline assessment was 50 years (sd 12.1; median 53 years, range 49). The mean disease duration at baseline was 10 years (sd 8.4; median 8 years, range 29). The mean interval between radiographs was 25 months (sd 9.6; median 26 months, range 36). The mean baseline score for the Steinbrocker, Ratingen, modified Sharp and Sharp-van der Heijde was 15.4 (sd 21.63), 13.2 (sd 25.23), 26.3 (sd 39.05) and 26.8 (sd 38.25).

The intra-rater reliability was high for all methods (Table 2). The baseline inter-rater reliability was high for the Sharp-van der Heijde, modified Sharp and Ratingen scores at 0.95 95% confidence interval (CI) 0.83-0.99, 0.94 (95% CI 0.78-0.96) and 0.89 (95% CI 0.64-0.97), respectively. The baseline inter-rater reliability for the Steinbrocker was low at 0.42 (95% CI -0.21-0.81). Review of the 10 films revealed the source to be poor agreement on the presence of periarticular osteopenia in three of the ten cases. All 50 baseline radiographs were therefore scored using the Steinbrocker by both readers to better reflect the frequency of osteopenia in PsA and hence more accurately estimate the performance of the Steinbrocker. The baseline inter-rater reliability for all 50 radiographs was 0.88 (95% CI 0.77-0.94).

The sensitivity to change of the methods is reported in Table 3. Using the smallest detectable change expressed as a percentage of the total score allows comparison between scores. The Sharp-van der Heijde has the greatest ability to detect change followed by the modified Sharp, Ratingen and Steinbrocker at 1.2%, 1.4%, 2.1% and 2.9% respectively. The sensitivity to change of the methods using the standardised

response mean demonstrated the Sharp-van der Heijde followed by the modified Sharp score to have the greatest ability to detect change at a level approaching 0.80. The Steinbrocker and Ratingen scores showed less sensitivity to change with levels of 0.46 and 0.44 respectively.

The feasibility of each method was estimated based on the mean time taken to score each film. The Steinbrocker took the least time to score followed by the Ratingen, Sharp-van der Heijde and modified Sharp at 6.2 minutes, 10.5 minutes, 14.4 minutes and 14.6 minutes respectively.

Discussion

Radiographic assessment of joint damage in psoriatic arthritis is an important outcome measure in longitudinal observational studies however the optimal method for use in this setting has not been determined and to our knowledge there are no reports comparing the existing methods. We report a comparison of four radiographic scoring methods used in psoriatic arthritis to inform discussion on the optimal method for use in longitudinal observational studies.

In terms of feasibility each of these methods are readily available and interpretable as a simple summative score. The time required to apply each method differs considerably in our study from 6.2 minutes to 14.6 minutes. An essential attribute of a scoring method for use in longitudinal observational studies is that it can be readily learned and applied feasibly. This 'feasibility barrier' has resulted in very limited radiographic data collection from observational cohorts. As may be expected, the global score (the Steinbrocker) is the most feasible and the time required to apply the composite scores is proportional to the number of areas scored (Ratingen<Sharp-van der Heijde<modified Sharp). Whilst there is no threshold of time beyond which a score becomes unfeasible, the Sharp-van der Heijde and modified Sharp scores are more challenging in terms of time required for application in longitudinal observational studies.

In our study we found good agreement between assessors and good test-re-test reliability. The exception was osteopenia in the Steinbrocker. The inter/intra-rater reliability estimates for this study (baseline inter 0.88 and intra for both readers 0.99/ 1.0) are comparable with the original reports (inter 0.86 and intra for both readers

0.81/ 0.80) [7]. We found the inter-rater reliability was poor (0.42) amongst the initial ten patients due to disagreement in three cases of possible osteopenia but rose to more acceptable levels (0.88) when applied to all 50 patients because less osteopenia was seen. The prevalence of periarticular osteopenia in psoriatic arthritis has not been reported but is thought to be less than rheumatoid arthritis[15]. Our group has previously reported a strong correlation between the modified Sharp score and periaticular osteopenia, present in 25 of the 73 psoriatic arthritis patient hand and feet radiographs included in the study[1]. Radiographic osteopenia is dependent on radiographic technique, varying according to choice of projection, exposure and capture media, thus may vary significantly between time-points and radiographers. Determining the presence of osteopenia is also subjective and therefore prone to variation. For these reasons osteopenia was removed from a number of radiographic scoring techniques in rheumatoid arthritis [16]. It may be argued therefore that any potential benefits of retaining osteopenia in a measure for use in longitudinal observational studies of psoriatic arthritis are outweighed by these disadvantages.

Regarding sensitivity to change we found the smallest detectable change to be greater than the mean change over two years in all techniques which is an important finding. The minimal detectable change/ smallest detectable change/ difference are all study specific (a function of mean change and measurement error) but are infrequently reported in trials [17]. There are no previous reports directly comparing the sensitivity to change of the Steinbroker, Sharp-van der Heijde and modified Sharp scores in psoriatic arthritis. The original Ratingen score method reported a minimal detectable change (calculated from the square root of the standard deviation of the inter-rater variance) for the total score, proliferation score and destruction score of 16.5, 8.4 and

11.5, respectively, which are comparable with our smallest detectable difference of 12.7, 5.8 and 7.3, respectively[8]. Guillemin *et al.* examined the reproducibility and sensitivity to change of five methods including the Sharp-van der Heijde in rheumatoid arthritis [18]. The mean change in the Sharp-van der Heijde score exceeded the smallest detectable difference for the total, erosion and joint space narrowing scores in their study (9.7, 5.8 and 7.2, respectively) yet the smallest detectable difference remains comparable with our study (10.8 , 7.3 and 7.3, respectively). Sharp *et al* examined the variability of precision of the modified Sharp method among readers from the datasets of six studies in rheumatoid arthritis [19]. Considerable variability between readers and the mean smallest detectable difference was greater than the mean progression, as we have found in our present study. We found that the smallest detectable difference/ change (normalised as a percentage of the total score to allow comparison) estimates were ranked as might be expected (Steinbrocker > Ratingen > modified Sharp > Sharp-van der Heijde). The smallest detectable difference/ change are largest for the global Steinbrocker score which assesses radiographic change as either present or absent rather than graded, thereby allowing less flexibility to detect change. The Sharp-van der Heijde is consistently the most sensitive to change than the score from which it was derived (the modified Sharp score). Finally the Ratingen score is more sensitive than the Steinbrocker as it is a composite rather than global score allowing grading of erosion and proliferation but not as sensitive as the Sharp-van der Heijde partly because it includes the wrist as a single joint rather than multiple small joints.

Other parameters within the scoring methods are worth noting. The soft tissue swelling element of the Steinbrocker is an additional source of variability, particularly

at the MTP joints there is a less clear view of the soft tissues. The scoring of joint space narrowing is not specific to psoriatic arthritis and can occur in concurrent osteoarthritis and thus over-estimate progression. The Ratingen score includes a measure of proliferation which was the only radiographic change sufficiently specific to psoriatic arthritis to justify inclusion in the CASPAR criteria [9]. Finally, the composite scores preserve data separately on erosion, joint space narrowing or in the case of the Ratingen score proliferation whereas such information is limited in a global score.

The findings of this study should be interpreted in light of certain methodological limitations. We have not blinded to the order of films and have both films available for comparison when scoring introducing the possibility of expectation bias and thus overestimating change. However, blinding to films may result in failure to detect progression. A study in rheumatoid arthritis has shown that blinding assessors to the chronology of films can introduce a measurement error that results in a loss of signal and hence underestimation of progression[20]. A second study from the same group showed that expert raters agreed most when the sequence of films was known[21]. Furthermore, two of the methods, the Ratingen score and Sharp-van der Heijde scores were both developed in un-blinded studies with known chronological order and so we have applied these tools as they were developed. We have selected patients with established disease in the two years prior to anti-TNF therapy and therefore likely to have progressive disease in order to study and compare the sensitivity to change of each method. The exact chronology of radiographic change in psoriatic arthritis is not yet established and therefore it is possible the scores may perform differently in early disease. Finally this study is informing the measurement of radiographic outcome in

observational studies where radiographs are most likely to be scored in known order and unblinded. A final potential limitation is that we have scored the radiographs with the assumption of no improvement as stipulated in the Sharp-van der Heijde scoring instructions. We applied the same rule to all methods to avoid biasing the results. This may have had the effect of overestimating the standardised response mean of the Steinbrocker which includes soft tissue swelling which may well improve over the study period.

We have found that the modified Sharp and Sharp-van der Heijde methods are the most reliable and sensitive to change in this present study, but took longer to perform. The Steinbrocker is the most feasible but lacks the sensitivity of the modified Sharp and Sharp-van der Heijde methods. The smallest detectable change of the Ratingen score is close to that of the modified Sharp and Sharp-van der Heijde but is quicker to perform and may be more specific to PsA through inclusion of proliferation. The findings of this study can be used to inform discussion on potential modifications and further study of these existing radiographic scoring methods for use in longitudinal observational studies of psoriatic arthritis.

Acknowledgements:

We would like to thank the patients for their participation in this research project. We would also like to thank Nicola Waldron our nurse specialist in psoriatic arthritis, Charlotte Cavill our database manager and Mandy Knight and Austin Smith our database administrators.

Funding:

This research is funded through an unrestricted grant from Abbvie laboratories ltd.

References

1. Ravindran J, Cavill C, Balakrishnan C, Jones SM, Korendowych E, McHugh NJ: A modified Sharp score demonstrates disease progression in established psoriatic arthritis. Arthritis Care Res (Hoboken) 2010, 62(1):86-91.
2. Gladman DD, Thavaneswaran A, Chandran V, Cook RJ: Do patients with psoriatic arthritis who present early fare better than those presenting later in the disease? Ann Rheum Dis 2011, 70(12):2152-2154.
3. Gladman DD, Mease PJ, Strand V, Healy P, Helliwell PS, Fitzgerald O *et al*: Consensus on a core set of domains for psoriatic arthritis. J Rheumatol 2007, 34(5):1167-1170.
4. Sharp JT, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL *et al*: Reproducibility of multiple-observer scoring of radiologic abnormalities in the hands and wrists of patients with rheumatoid arthritis. Arthritis Rheum 1985, 28(1):16-24.
5. van der Heijde D, Sharp J, Wassenberg S, Gladman DD: Psoriatic arthritis imaging: a review of scoring methods. Ann Rheum Dis 2005, 64 Suppl 2:ii61-64.
6. van der Heijde D: How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000, 27(1):261-263.
7. Rahman P, Gladman DD, Cook RJ, Zhou Y, Young G, Salonen D: Radiological assessment in psoriatic arthritis. British journal of rheumatology 1998, 37(7):760-765.
8. Wassenberg S, Fischer-Kahle V, Herborn G, Rau R: A method to score radiographic change in psoriatic arthritis. Z Rheumatol 2001, 60(3):156-166.
9. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H: Classification criteria for psoriatic arthritis: development of new criteria from a large international study. Arthritis Rheum 2006, 54(8):2665-2673.
10. Tillett W, Costa L, Jadon D, Wallis D, Cavill C, McHugh J *et al*: The ClASsification for Psoriatic ARthritis (CASPAR) criteria--a retrospective feasibility, sensitivity, and specificity study. J Rheumatol 2012, 39(1):154-156.
11. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, vad de Putte LB: Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989, 1(8646):1036-1038.
12. Taylor WJ, Porter GG, Helliwell PS: Operational definitions and observer reliability of the plain radiographic features of psoriatic arthritis. J Rheumatol 2003, 30(12):2645-2658.
13. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986, 1(8476):307-310.
14. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D: Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005, 64(2):179-182.
15. Ory PA, Gladman DD, Mease PJ: Psoriatic arthritis and imaging. Ann Rheum Dis 2005, 64 Suppl 2:ii55-57.
16. Boini S, Guillemin F: Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann Rheum Dis 2001, 60(9):817-827.

17. van der Heijde D, Landewe R, Klareskog L, Rodriguez-Valverde V, Settas L, Pedersen R *et al*: Presentation and analysis of data on radiographic outcome in clinical trials: experience from the TEMPO study. Arthritis Rheum 2005, 52(1):49-60.
18. Guillemin F, Billot L, Boini S, Gerard N, Odegaard S, Kvien TK: Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. J Rheumatol 2005, 32(5):778-786.
19. Sharp JT, Wolfe F, Lassere M, Boers M, Van Der Heijde D, Larsen A *et al*: Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. J Rheumatol 2004, 31(6):1062-1072.
20. van Der Heijde D, Boonen A, Boers M, Kostense P, van Der Linden S: Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology (Oxford) 1999, 38(12):1213-1220.
21. Bruynesteyn K, van der Heijde D, Boers M, Lassere M, Boonen A, Edmonds J *et al*: Minimal clinically important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts. J Rheumatol 2001, 28(4):904-910.

Table 1. Summary of scoring methods

Scoring methods	Scales	Total erosion score	Total JSN score	Total score
STB	**42 Joints of the hands and feet: Scale 0-4** 0 is normal. 1 is juxta-articular osteopenia or soft tissue swelling. 2 indicates the presence of any erosion. 3 the presence of erosion and joint space narrowing. 4 total joint destruction (lysis or ankylosis).	N/A	N/A	168
MSS	**54 joints (42 hands, 12 feet) for erosion: scale 0-5** 0 = no erosion. 1 = one discrete erosion or involvement of < 21% of the joint are by erosion. 2 = two discrete erosions or involvement of 21- 40% of the joint. 3 = three discrete erosions or involvement of 41-60% of the joint. 4 = four discrete erosions or involvement of 61-80% of the joint. 5 = extensive destruction involving more than 80% of the joint. **54 joints (44 hands, 10 feet) for JSN: scale 0-4** 0 = normal joint. 1 = asymmetrical or minimal narrowing. 2 = definite narrowing with loss of up to 50% for the normal space. 3 = definite narrowing with loss of 51 – 99% of the normal space. 4 = absence of a joint space, presumptive evidence of ankylosis. 5 = widening.	270	216	486
VDH	**52 joints (42 hands, 10 feet) for erosion: scale 0-5 (hands) 0-10 (feet)** 0 = no erosions. 1= discrete erosion. 2 = large erosion not passing the midline. 3 = large erosion passing the mid-line. 4 = combination of below 5 = combination of below **52 joints (42 hands, 10 feet) for JSN: scale 0-4** 0 = normal 1 = asymmetrical minimal narrowing with loss of up to a maximum of 25% 2 = definite narrowing with loss of up to 50% of the normal space 3 = definite narrowing with loss of 50 – 99% of the normal space or subluxation 4 = absence of a joint space, presumptive evidence of ankylosis, or complete subluxation.	320	208	528
PARS	**40 joints (30 hands, 10 feet) for destruction: scale 0-5** 0 = normal. 1 = one or more definite erosion with an interruption of the cortical plate of >1mm but destruction of < 10% of the total joint surface. 2 = destruction of 11 – 25%. 3 = destruction of 26 – 50%. 4 = destruction of 51 – 75% . 5 = destruction of more than 75% of the joint surface. **40 joints (30 hands, 10 feet) for proliferation: scale 0-4** 0 = normal. 1 = bony proliferation measured from the original bone surface of 1-2mm, or clearly identifiable bone growth not exceeding 25% of the original diameter of the bone. 2 = bony proliferation of 2-3mm or bone growth between 25 to 50% 3 = bony proliferation >3mm or bone growth >50% 4 = bony ankylosis.	200	160 *	360

Steinbrocker method (STB), Sharp van der Heijde method (VDH), Modified Sharp method (MSS), Ratingen method (PARS), Joint space narrowing (JSN) *Proliferation

Table 2. Inter/ Intra-rater reliability of each scoring method

Method		Range	Inter rater reliability (95%CI)		Intra rater reliability (95%CI)	
			Baseline	Follow up	Rater 1	Rater 2
STB		0-168	0.42 (-0.21-0.81)	0.40 (-0.23-0.81)	0.99 (0.95-1.0)	1.00 (0.99-1.00)
	STB (n=50)		0.88 (0.77-0.94)			
MSS		0-486	0.94 (0.78-0.96)	0.96 (0.86-0.99)	0.99 (0.95-1.00)	1.00 (0.99-1.00)
	Erosion	0-270	0.77 (0.35-0.94)	0.64 (0.10-0.90)	0.77 (0.35-0.94)	1.00 (0.99-1.00)
	JSN	0-216	0.96 (0.86-0.99)	0.94 (0.81-0.99)	0.95 (0.81-0.99)	0.99 (0.98-1.00)
VDH		0-528	0.95 (0.83-0.99)	0.99 (0.96-1.00)	0.97 (0.90-0.99)	0.99 (0.98-0.99)
	Erosion	0-320	0.91 (0.70-0.98)	0.92 (0.72-0.98)	0.91 (0.69-0.98)	1.00 (0.99-1.00)
	JSN	0-208	0.96 (0.87-0.99)	0.92 (0.73-0.98)	0.93 (0.76-0.98)	0.98 (0.97-0.99)
PARS		0-360	0.89 (0.64-0.97)	0.90 (0.65-0.97)	0.99 (0.95-1.00)	0.99 (0.97-1.00)
	Destruction	0-200	0.69 (0.18-0.91)	0.69 (0.18-0.91)	0.76 (0.31-0.93)	1.00 (0.99-1.00)
	Proliferation	0-160	0.90 (0.67-0.97)	0.85 (0.52-0.96)	0.90 (0.66-0.97)	1.00 (0.98-1.00)

Modified Steinbrocker method (STB), Sharp-van der Heijde modified method (VDH), Modified Sharp method (MSS), psoriatic arthritis Ratingen score (PARS), Joint space narrowing (JSN)

Table 3. Sensitivity to change of each scoring method

Method		Mean Change	SD of change	SEM	SRM	SDD	SDC	SDD as % of total score	SDC as % of total score
STB		2.3	4.91	3.49	0.46	8.11	4.83	4.82	2.87
PARS		3.3	7.61	5.46	0.44	12.71	7.57	3.53	2.10
	Destruction	1.6	3.46	2.51	0.45	5.83	3.48		
	Proliferation	1.8	4.28	3.15	0.43	7.34	4.37		
MSS		5.5	7.15	5.06	0.77	11.77	7.01	2.42	1.44
	Erosion	2.4	4.15	2.97	0.57	6.90	4.11		
	JSN	3.3	5.10	3.64	0.64	8.47	5.05		
VDH		5.2	6.53	4.66	0.79	10.83	6.45	2.05	1.22
	Erosion	2.3	4.41	3.14	0.52	7.31	4.36		
	JSN	3.0	4.36	3.14	0.68	7.29	4.35		

Modified Steinbrocker method (STB), Sharp-van der Heijde modified method (VDH), Modified Sharp method (MSS), psoriatic arthritis Ratingen score (PARS), Joint space narrowing (JSN), Standard deviation (SD), Standard error of means (SEM), Standardised response mean (SRM), Smallest detectable difference (SDD), Smallest detectable change (SDC).